

## ΣΥΣΤΗΜΑΤΑ και ΤΕΧΝΟΛΟΓΙΕΣ ΓΝΩΣΗΣ

### ΑΣΚΗΣΗ 3<sup>η</sup>

Προαιρετική - Ημερομηνία παράδοσης: τέλος εαρινής εξεταστικής περιόδου

## Μηχανική Μετάφραση Αριθμητικών

### Αντικείμενο - στόχος

Θέλουμε να μεταφράζουμε μηχανικά την ολογραφική μορφή των αριθμητικών από μια φυσική γλώσσα σε μια άλλη.

*Περιορισμός περιβάλλοντος:* Περιοριζόμαστε στα απόλυτα αριθμητικά (π.χ. «δέκα τρία») ουδετέρου γένους (π.χ. «δέκα τρία») και δεν εξετάζουμε τα άλλα γένη (π.χ. «δέκα τρεις») ούτε τα τακτικά αριθμητικά (π.χ. «δέκατο τρίτο»).

### Παραδείγματα αριθμητικών σε τέσσερις γλώσσες:

Ελληνικά: εκατό, ενενήντα τρία, εξακόσια δώδεκα, ...  
Αγγλικά: a hundred, ninety three, six hundred twelve, ...  
Γαλλικά: cent, quatre vingt treize, six cents douze, ...  
Γερμανικά: ein Hundert, drei und neunzig, sechs Hundert zwei, ...

### Γλώσσα

Επιλέξτε γλωσσικό ζευγάρι: Την Ελληνική γλώσσα και μία από τις γλώσσες Αγγλική, Γαλλική ή Γερμανική.

### Φορμαλισμός – Υπολογιστικό περιβάλλον

Χρησιμοποιείτε φορμαλισμό Ενοποιητικής Γραμματικής PATR II.

Μπορείτε να χρησιμοποιήσετε είτε το PCPatr είτε την υλοποίηση του Patr II σε Prolog. (Δείτε το «εκπαιδευτικό υλικό» στην ιστοσελίδα του μαθήματος).

### Σύστημα - Συντακτικός Αναλυτής (parser) Φυσικής Γλώσσας

Σχεδιάστε ένα σύστημα «μηχανικού μεταφραστή» (φυσικής γλώσσας), ο οποίος να δέχεται ένα αριθμητικό, από μηδέν μέχρι και εννιακόσια ενενήντα εννέα, σε μία γλώσσα (A) και να επιστρέφει το αριθμητικό αυτό μαζί με τη μετάφρασή του σε μια άλλη γλώσσα (B). Η υλοποίησή σας θα πρέπει να λειτουργεί και για την ανάστροφη μετάφραση (από τη γλώσσα B στη γλώσσα A).

Περιγράφουμε μια γλώσσα (απόλυτων) αριθμητικών, όπου 'νόμιμες προτάσεις' είναι τα αριθμητικά που αντιστοιχούν στους μονοψήφιους, διψήφιους ή τριψήφιους αριθμούς (σε ολογραφική μορφή ή ολογράφως). Οι νόμιμες προτάσεις συγκροτούνται από μία ή περισσότερες λέξεις της φυσικής γλώσσας.

Π.χ. (στην ελληνική γλώσσα): μηδέν, ένα, δύο, τρία, ... δέκα, έντεκα, δώδεκα, ... δέκα πέντε, ... είκοσι ένα, ... εκατό, ... εκατόν πέντε, ... διακόσια δώδεκα. ... κ.λπ. μέχρι εννιακόσια ενενήντα εννέα.

Κάθε τέτοια 'νόμιμη πρόταση' της φυσικής γλώσσας ελέγχεται ως προς την ορθότητά της από το σύστημα και δημιουργείται μια δομή (ενοποιητικής γραμματικής) η οποία αντλεί τις ιδιότητες και τις τιμές κάθε 'λέξης' από το 'λεξικό'.

Γιαυτό, αρχικά, καταγράφουμε όλες τις ιδιομορφίες που παρουσιάζει το ζευγάρι γλωσσών που επιλέξαμε και οργανώνουμε το 'λεξικό' κατάλληλα ώστε να επιλέγεται για κάθε νόμιμη δομή το κατάλληλο λεκτικό, μέσω ιδιοτήτων και τιμών (paths: attribute-value pairs) των 'φραστικών συστατικών' του, εν προκειμένω των λέξεων.

### *Υπόδειξη: Ιδιαιτερότητες Αριθμητικών*

Αμφισημία:

Το απόλυτο αριθμητικό που αντιστοιχεί στο 100 γράφεται αλλιώς («εκατό») αν δεν έχει παρακολουθήμα και αλλιώς («εκατόν ...») αν ακολουθείται από άλλον αριθμό (π.χ. «εκατόν πέντε»).

Στις διαφορετικές γλώσσες η μία λέξη μεταφράζεται σε δύο ή τρεις και αντιστρόφως:

εκατό => a hundred, ein Hundert  
treize, quatorze, quinze ... => δέκα τρία, δέκα τέσσερα, δέκα πέντε  
thirteen, fourteen, fifteen ... => δέκα τρία, δέκα τέσσερα, δέκα πέντε  
ογδόντα => quatre vingt  
ενενήντα => quatre vingt dix  
εξακόσια => six hundred